Karlsson on databases and stuff

I am Anders Karlsson, and I have been working in the RDBMS industry for many, possibly too many, years. In this blog, I write about my thoughts on RDBMS technology, happenings and industry, and also on any wild ideas around that I might think up after a few beers.

Friday, June 1, 2012

JSON to MySQL - mysqljsonload 1.0 Ready for download!

I have spent some free time developing this tool that loads JSON data into MySQL. And yes, I know there are some other means of doing this, but in my case, yes, this was done with a particular purpose in mind, I needed something different. Something that could load a lot of JSON data real fast.

The program is a C program, it uses the Jansson library for JSON parsing (I really like this library by the way. It's available here: http://www.digip.org/jansson/) and it is multi threaded (configurable number of loading threads) and uses MySQL INSERT arrays.

You want to try it? It's available on sourceforge download. The program comes with some basic documentation, which is also available as a separate download, and uses the usual configure / make process to build it.

There are a bunch of configuration options, as usual these can be either in a configuration file or on the command line, to support things like:

- Ignoring specific columns.
- Mapping column names, i.e. in the normal case attributes in the JSON data is mapped to a database column with that same name, but this allows you to change that.
- Set attribute defaults. If an attribute doesn't exist in JSON, by default it is set to NULL, but you may set it to some other value using this option.
- Column fixed values. Sometimes you want to assign a column a fixed value, independent of what it is set to in the JSON file or if it is set att all. That is what this option does for you.
- NULL handling of embedded objects and arrays. In the normal case, objects and array embedded in the JSON data records get loaded as string, but you may turn this off and load them as NULL instead.

There is a bunch more things this program can do, but this is the basic stuff. I tested the program by loading a 23 Gb JSON file containing about 100.000.000 records into an InnoDB table. The loading went ahead with about 10500 records per second, which I think is reasonable on this mid-size machine (1 CPU, 8 cores, 16 Gb RAM, MySQL running on a single 1Tb disk, the same as I was loading data from). MySQL was configured with InnoDB lazy logwrites and a few other things, but nothing fancy.

So, if you are still with me, you might have a reason for that. Like a MongoDB export file (with JSON data) that you want to cram into a MySQL database? This this tool may be for you.

Also, this is version 1.0. There is probably a whole bunch of bugs, some not so nice code that I want to clean up and some obvious features to add. Let me know what you think! And if you want a quick introduction to JSON, look at my previous post on that subject.

/Karlsson
Yes, that Karlsson not Json

Posted by Anders Karlsson at 3:02 PM

1 comment:

About Me

Anders Karlsson

View my complete profile

Blog Archive

- ► 2014 (8)
- ► 2013 (26)
- ▼ 2012 (52)
  - ► December (1)
  - ► November (3)
  - ► October (3)
  - ► September (2)
  - ► August (12)
  - ► July (7)
  - ▼ June (6)
    - JSON Loader version 1.1 available!
    - Is query_cache really a good idea? Maybe not...
    - A little something on the basics of UTF-8
    - No MySQL Cluster, the table isn't bl**dy full! Not...
    - Setting up MySQL Cluster 7.2
    - JSON to MySQL - mysqljsonload 1.0 Ready for downlo...
  - ► May (4)
  - ► April (3)
  - ► March (2)
  - ► February (7)
  - ► January (2)
- ► 2011 (54)
- ► 2010 (49)
- ► 2009 (55)

John Lynn said...

Great idea! I need this now, but the sourceforge download seems to only give me the .pdf file.

January 31, 2013 at 7:33 PM

Post a Comment

► 2008 (4)

Newer Post Home Older Post

Subscribe to: Post Comments (Atom)

Watermark template. Powered by Blogger.